Stopping executions, saving computers with new malware detection tool

October 21, 2009

Virus detection takes a smarter turn with information-based approach

Los Alamos, New Mexico, October 13, 2009—In the cyber security world, looking for viruses or other malicious files on computers opens the door to potentially releasing the file and contaminating one's system. In classical terms, if you peer into Medusa's eyes, you're turned to stone. Greek hero Perseus turned a mirror on the monster to protect himself, and in similar fashion a team of scientists from LANL's International, Space, and Response Division and LANL's Advanced Computing Laboratory has patented a computer tool that allows the machine to identify malicious executable files without being exposed to their harmful actions. Malicious executables, often spread as e-mail attachments, pose serious security threats to computer systems and associated networks.

"This tool would help to advance the practice of cyber security from an expert-based approach to an information-based approach," said Michael Cai of LANL's International, Space, and Response Division, who along with James Theiler of the same division and Maya Ghokale of LANL's Advanced Computing Laboratory, developed the new method. As with many LANL technologies that could be useful in the commercial world, "the Laboratory seeks partners to commercialize this technology," said David Seigel of the LANL Technology Transfer Division "and there may be opportunities for exclusive field-of-use licensing," he noted.

The team determined that two main categories of information were critical to catch the problem programs before they could act: Features, or the most informative characteristics that describe the programs, plus Classifiers, in this case "support vector machine classifiers" (SVM), which apply the selected features to identify those suspicious programs in a reliable and optimal way. This is the first patent to use the SVM technique for computer virus detection, and it is able to detect known, novel, and unknown viruses.

"This new technology is another weapon against cyber threats, one that is not easily spoofed or confused," said Theiler. "This tool could make computer systems more resilient and efficient against the onslaught of malicious software that comes their way."

The team investigated the use of byte-sequence frequencies as a way to automatically distinguish malicious from benign executables without executing them. Byte sequence frequencies are occurrences of each character, for example, the letter "A," or the

space, in a particular program. The typical telltale sign is that malicious programs have different distributions of byte-sequence frequencies from those that are benign. The advantage for choosing byte sequences as features is that those byte patterns are the most accessible, and at the same time, the most direct information in a program. In real-world terms, this tool could be used in antivirus software for institutional users, and for anomaly detections.

The work was supported by the Laboratory-Directed Research and Development (LDRD) program at Los Alamos. Those interested in more information about licensing opportunities for this technology can contact David Seigel at seigel@lanl.gov.